

pp. 325–326:

Minimax diameter clustering.

Form of partitioned clustering.

Want to minimize the maximum diameter of any cluster.

Farthest-point clustering for any metric space [T. Gonzales,

Clustering to minimize the maximum intracluster distance,

Theoretical Computer Science, 38:293–306, 1985]:

start with an arbitrary point s_1 .

Pick a point s_2 that is as far from s_1 as possible.

Pick s_i to maximize the distance to the nearest of

all centroids picked so far.

That is: maximize the $\min\{\text{dist}(s_i, s_1), \text{dist}(s_i, s_2), \dots\}$.

After all k representatives are chosen we can define

the partition of S : cluster S_j consists of all points closer to

s_j than to any other representative.

Theorem: Let the maximum diameter of this k clustering C_{gon}

be D , then even in the optimal k clustering C_{opt}

the maximum diameter must be at least $D/2$.

Here optimal is in the sense of minimizing the maximum diameter of any cluster.

Proof: C_{gon} has centroids s_1, \dots, s_k .

Add one more point q using farthest-point clustering.

Let E be the minimum distance from q to any s_1, \dots, s_k .

Now, by construction, all the other centroids must be at least

E apart from one another.

Therefore, if we are creating an

optimal k -clustering, because we must include these $k+1$ points,

some cluster X must have two of these points and those

two points must be at least E apart.

So, the lower bound on the optimal clustering is a diameter of E .

Now, consider C_{gon} . Point q must be in the cluster Y of some centroid

s_i , so that cluster must have radius at most E and hence diameter

at most $2E$.

No other point in any cluster of C_{gon} can be farther from

its centroid by construction of q .

Fast implementation: clusters points into rectangular boxes

and then simply compute from the rectangle.

Remarkably there is a further theorem that says that it is NP-hard

to do any better than a factor of 1.962 (this result is only for

Euclidean).